Samantha Gregoryk
November 9th, 2020
Ananth Jillepalli
CPTS 315
Homework 3

Q1. Answer the following with a yes or no along with proper justification.

a. Is the decision boundary of voted perceptron linear?

   No, because there are nonlinear boundaries. Voted perceptron can leave boundaries in the positive and negative quadrants and therefore the linear pattern is no longer applicable.

b. Is the decision boundary of averaged perceptron linear?

   Yes, because there are two weights that are used to find the averaged perceptron and that causes a linear decision boundary.

Q2. Consider the following setting. You are provided with n training examples: (x_1, y_1, h_1), (x_2, y_2, h_2), …, (x_n, y_n, h_n) where x_i is the input example, y_i is the class label (+1 or -1), and h_i > 0 is the importance weight of the example. The teacher gave you some additional information by specifying the importance of each training example.

- How will you modify the perceptron algorithm to be able to leverage this extra information? Please justify your answer.

  If we give a high weight to a training data, it will put an emphasis on the training data. I would modify the perceptron algorithm by using voted perceptron. This will help keep those results more accurate and will help with over-fitting. We would do this by multiplying the weight of the training data set for every perceptron algorithm pass to get the modified loss.

Q3. Consider the following setting. You are provided with n training examples: (x_1, y_1), (x_2, y_2), … , (x_n, y_n), where x_i is the input example, and y_i is the class label (+1 or -1). However, the training data is highly imbalanced (say 90% of the examples are negative and 10% of the examples are positive) and we care more about the accuracy of positive examples.

- How will you modify the perceptron algorithm to solve this learning problem? Please justify your answer.

  Since there is not as much training data used as the previous example and the examples are off-balanced, it is better to modify the perceptron algorithm by using geometric margins because the results would be more linearly separable. It would also be better to modify the perceptron algorithm by using maximum margin classifiers because we can find the linear classifier with a maximum margin.

Q4. You were just hired by MetaMind. MetaMind is expanding rapidly, and you decide to use your machine learning skills to assist them in their attempts to hire the best. To do so, you have the following available to you for each candidate *i* in the pool of candidates *I:*

(i) Their GPA
(ii) Whether they took Data Mining course and achieved an A
(iii) Whether they took Algorithms course and achieved an A
(iv) Whether they have a job offer from Google
(v) Whether they have a job offer from Facebook
(vi) The number of misspelled words on their resume

You decide to represent each candidate $i \in I$ by a corresponding 6-dimensional feature vector $f(x^i)$. You believe that if you just knew the right weight vector $w \in R^6$ you could reliably predict the quality of a candidate *i* by computing $w \cdot f(x^i)$. To determine *w* your boss lets you sample pairs of candidates from the pool. For a pair of candidates *(k, l)* you can have them face off in a "DataMining-fight." The result is score *(k > l),* which tells you that candidate k is at least score *(k > l)* better than candidate *l*. Note that the score will be negative when *l* is a better candidate than *k*. Assume you collected scores for a set of pairs of candidates *P*.

a. Describe how you could use a perceptron-based algorithm to learn the weight vector w. Describe the basic intuition. How the weight updates will be done?

   If *score(k > l) = - score(k > l)* & *score (k > l) != 0* for any pair $(k, l) \in P$, we could use margin based perceptron because we have a two candidate pairs *k* ((w, $f(x^k)$)) and *l* ((w, $f(x^l)$)) where we want to compare if *(w,* $f(x^k)$*)* ≥ *(w,* $f(x^l)$*) + score(k > l)* (if score(k > l) > 0) OR if *(w,* $f(x^l)$*)* ≤ *(w,* $f(x^k)$*) - score(k > l)* is equal to *(w,* $f(x^k)$*) + score(l > k).*

b. Use Pseudo-code for the entire algorithm.

```
MarginPerceptron() {
        random_pair = get a random pair (k, l) from the candidates P

        if the score(k > l) > 0:
                if (w, $f(x^k)$) ≥ (w, $f(x^l)$) + score(k > l):
                        return nothing
                else:
                        update w by using w + $f(x^l)$ - $f(x^k)$
        if the score(k > l) < 0:
                if (w, $f(x^k)$) ≥ (w, $f(x^l)$) - score(k > l):
                        return nothing
                else:
                        update w by using w + $f(x^l)$ - $f(x^k)$
}
```

Q5. Suppose we have n_+ positive training examples and n_- negative training examples. Let C_+ be the center of the positive examples and C_- be the center of the negative examples,

$$\text{i.e. } C_+ = \frac{1}{n_+}\sum\nolimits_{i:\, y_i = +1} x_i \text{ and } C_- = \frac{1}{n_-}\sum\nolimits_{i:\, y_i = -1} x_i$$

Consider a simple classifier called CLOSE that classifies a test example x by assigning it to the class whose center is closest.

a. Show that the decision boundary of the CLOSE classifier is a linear hyperplane of the form *sign(w · x + b).* Compute the values of *w* and *b* in terms of C_+ and C_-

$$\|x - C_+\| \leq \|x - C_-\|$$

$$\|x - C_+\|^2 \leq \|x - C_-\|^2$$

$$\|x\|^2 - 2xC_+ + \|C_+\|^2 \leq \|x\|^2 - 2xC_- + \|C_-\|^2$$

$$2xC_+ + \|C_+\|^2 \leq 2xC_- + \|C_-\|^2$$

$$2xC_+ - 2xC_- + \|C_-\|^2 - \|C_+\|^2 \geq 0$$

$$w = 2x(C_+ - C_-)$$

$$b = \|C_-\|^2 - \|C_+\|^2$$

Recall that the weight vector can be written as a linear combination of all the training examples:

$$w = \sum_{i=1}^{n_+ + n_-} a_i \cdot y_i \cdot x_i$$

b. Compute the dual weights (*a's*).

$$w = 2\sum_{i=1}^{+n_+} x_i - 2\sum_{i=1}^{-n_-} x_j$$

$$n_+ = \frac{2}{n_+}$$

$$n_- = \frac{2}{n_-}$$

c. How many of the training examples are support vectors?

Out of all the training examples, the one that are support vectors are the ones closest to the linear separator line.

Q6. (20 points) Please read the following paper and write a brief summary of the main points in at most TWO pages.

Pedro M. Domingos: A few useful things to know about machine learning. Communications of ACM 55(10): 78-87 (2012) https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

With the mass amounts of data, we see receive on a daily basis, machine learning is more advanced than ever before. Learning the processes of machine learning algorithms is vital for a company's success. The companies that invest in time and resources to make machine learning a key feature in their programs are the ones that thrive in the long run. Experimenting with different learners, data sources, and learning problems will only lead to improvements, but knowing how to approach these ideas with the data you hold can make or brake systems.

Mistakes are commonly made when testing on the training data. Sometimes, if the chosen classifier is then tested on new data, it is often no better than random guessing. Depending on the data given, it may not always be the best set for that algorithm. As said in the article, if you are hiring someone to build a classifier, be sure to keep some of the data to yourself and test the classifier they give you on it. This will ensure parts of your data work and can be tested on the entire data set later for accurate results.

Also, as a rule of thumb, a dumb algorithm with lots and lots of data beats a clever one with modest amounts of it. Let the machine learning algorithms do the heavy lifting to get the best results needed for success. Be cautious about the scalability of your algorithms. Time, memory, and for specifically machine learning, training data can cause major errors if not treated properly.

The goal of machine learning is to generalize beyond the examples in the training set. Having a conclusion from the results of your data might not be enough. Intuition plays a big role in machine learning because the algorithms can only do as much as we tell them to. Being able to create assumptions beyond the data that was given is important to generalize beyond it. Use these results as “guides of action”.